

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: December 9, 2002, 12:24:14 ; Search time 2054 Seconds  
(without alignments)  
2633.541 Million cell updates/sec

Title: US-09-765-034-2  
Perfect score: 1747  
Sequence: 1 MGIIMANNATCKNMLAENA.....KSLTFSRMAHELLLSPREK 334

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cg2.1/USPTO.spool/US09765034/runat.05122002.132016.21489/app.query.fasta.1.519  
-DB=EST -QFWT=fastlap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human0.cdi -LIST=45  
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09765034.ecgn.1.1.761/runat.05122002.132016.21489 -NCPU=6 -ICPU=3  
-NO.XLPYX -NO.MMAP -LARGOQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=10 -HRRADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlm:\*  
4: em\_estlm:\*  
5: em\_estlm:\*  
6: em\_estlm:\*  
7: em\_estlm:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_tod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	842.5	48.2	877	12	BG402029
2	737	42.2	633	17	AG083174 Pan trogl
3	692	39.6	1101	17	CNS04VY3
4	655	37.5	956	17	CNS028Y4
5	619	35.4	520	9	AL165305
6	536	30.7	1026	17	CNS051MY
7	500	28.6	422	10	BB847918
8	480	27.5	420	10	BB864882
9	478	27.4	589	12	BF196066
10	477	27.3	639	10	AM612141
11	467	26.7	582	10	BE221739
12	426	24.4	636	10	BB645274
13	402	23.0	416	10	BB846608
14	383	21.9	3001	11	AK017378
15	383	21.9	3001	11	AK005013
16	350.5	20.1	1258	11	AF345567
17	342	19.6	1022	17	CNS04W90
18	330	18.9	946	9	AL551903
19	328	18.8	931	9	AL547762
20	327	18.7	1766	11	AK008952
21	325	18.6	904	9	AL532537
22	322.5	18.5	623	14	B0038875
23	322.5	18.5	2146	11	BC027965
24	319.5	18.3	989	17	CNS03C14
25	316	18.1	2020	11	BC013202
26	306.5	17.5	3005	11	AK019478
27	300.5	17.2	1848	11	AK011967
28	289	17.1	2706	11	AK020001
29	296	16.9	641	14	B0396255
30	296	16.7	870	9	AL545172
31	292	16.7	885	9	AL546894
32	292	16.7	885	9	AL546894
33	286.5	16.4	1057	10	BM918711
34	285.5	16.3	1009	10	BB609892
35	285	16.3	907	9	AL552677
36	284	16.3	1017	14	BM922807
37	283.5	16.2	1063	14	BM920158
38	280.5	16.1	937	17	CNS03GSG
39	280.5	16.1	1372	11	AF345568
40	280	16.0	1074	14	BM917063
41	278	15.9	935	9	AL531903
42	276.5	15.8	942	17	CNS0455H
43	275	15.7	861	13	BI768868
44	274.5	15.7	1076	14	BM923028
45	272.5	15.6	1088	13	BM549799

## ALIGNMENTS

RESULT 1  
LOCUS BG402029 877 bp mRNA linear EST 12-MAR-2001  
DEFINITION 602466748F1 NIH\_MGC\_75 Homo sapiens cDNA clone IMAGE:4594810 5',  
mRNA sequence.  
ACCESSION BG402029  
VERSION BG402029.1 GI:13295477  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 877)  
NIH-MGC http://mgc.ncl.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE

JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: c9apds-remail.nih.gov  
 Tissue Procurement: CLONTECH Laboratories, Inc.  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LICM136 row: p column: 11  
 High quality sequence stop: 542.  
 Location/Qualifiers

FEATURES  
 source  
 1..877  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4594810"  
 /clone\_1lb="NIH MGC 75"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: Kidney; Vector: pDNR-LIB (Clontech); Site\_1:  
 SfiI (ggccgcctcgcc); Site\_2: SfiI (ggccatattggcc); 5' and  
 3' adaptors were used in cloning as follows: 5' adaptor  
 sequence: 5'-CACGGCCATTATGGC-3' and 3' adaptor sequence:  
 5'-ATCTAGAGCGCGCGCGCCAGACAG-dt(30)BN-3' (where B = A,  
 C, or G and N = A, C, G, or T). Average insert size 1.65  
 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts  
 by PCR. This library was enriched for full-length clones  
 and was constructed by Clontech Laboratories (Palo Alto,  
 CA). Note: this is a NIH-MGC Library."

BASE COUNT 198 a 221 c 199 g 259 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 5.06e-84 Length: 877  
 Score: 842.50 Matches: 176  
 Percent Similarity: 90.50% Conservative: 5  
 Best Local Similarity: 88.00% Mismatches: 16  
 Query Match: 48.23% Indels: 5  
 DB: 12 Gaps: 2

US-09-765-034-2 (1-334) x BG402029 (1-877)

OY 1 MetleuGlYIleMetAlaTrpAsnAlaThrCysIysAsnTrpLeuAlaAlaGluAla 20  
 DB 101 ATGCTGGGATCATGGATGGAATGCAACTGCAAACTGGCTGGCAGAGGCTGCC 160  
 OY 21 LeuGluIysTYrTYrLeuSerIlePheTYrGlyIleGluPheAlaValGlyValLeuGly 40  
 DB 161 CTGGAAGAAGTACTACCTTTCATTTTATGAGTATGAGTTCGTTTGCGAGCTCTTGA 220  
 OY 41 AsnThrIleValValTYrGlyTYrIlePheSerLeuIysAsnTrpAsnSerSerAsnIle 60  
 DB 221 AATACCATTTGTTTACGCTACATCTTCTCTGAAGAACTGGACAGCAGTAATATT 280  
 OY 61 TYrLeuPheAsnLeuSerValSerAspLeuAlaPheLeuCysThrLeuProMetLeuIle 80  
 DB 281 TATCTCTTAACCTCTGCTGCTGACTTACCTTTCGTGCACCCCTCCATGCTGATTA 340  
 OY 81 ArgSerTYrAlaAsnGlyAsnTrpIleTYrGlyAspValLeuCysIleSerAsnArgTYr 100  
 DB 341 AGGAGTATATGCAATGCAAACTGATATATGAGACGCTCTGCTATACGACACCGAAT 400  
 OY 101 ValLeuHisAlaAsnLeuTYrThrSerIleLeuPheLeuThrPheIleSerIleAspArg 120  
 DB 401 GTGCTTCATGCAACCTATATACAGCATTCCTTTCACATTTTATACAGATGATGATGA 460  
 OY 121 TYrLeuIleIleIysTYrTrpPheArgGluHisLeuLeuGlnIysGluPheAlaIle 140  
 DB 461 TACTGTAAATTAAGTATCCCTTTCGAGAACACCTTCGCAAAAGAAAGATTGCTATT 520  
 OY 141 LeuIleSerLeuAlaIleTYrValLeuValThrLeuGluLeuLeuProIleLeuProLeu 160  
 DB 521 AATCTC---CTGGCCATTGGGT-TTAGTAACTTAGAGTTACTACCATATCTTCCCTT 576

OY 161 ILeAsnProValIle-ThrAspAsnGlyThrThrCysAsnAspPheAlaSerSerGlyAs 180  
 DB 577 ATAAATCTGTATTAATGACAAATGGCGCC--TGTAAGATTTTGAAGGCTGGAGA 634  
 OY 180 ProAsnTYrAsnLeuIleTYrSerMetCysIleuThrLeuLeuGlyPheLeuIlePro 199  
 DB 635 CCCCAATCCCAACCTC---TTTTCAGATGTGTCTACCTGTGGGCGCTTATCC 689

RESULT 2  
 AG083174/c 633 bp DNA linear GSS 03-NOV-2001  
 LOCUS Pan troglodytes DNA, clone: PTB-080115.R, genomic survey sequence.  
 DEFINITION AG083174  
 ACCESSION AG083174.1 GI:16634976  
 VERSION  
 KEYWORDS  
 SOURCE

## ORGANISM

Pan troglodytes male lymphoblast DNA, clone\_1lb:PTB Chimpanzee Male  
 BAC library clone:PTB-080115.R.  
 Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

## REFERENCE

1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
 Totoki, Y., Watanabe, H. and Sakaki, Y.  
 BAC end sequences of library PTB

Unpublished  
 2 (bases 1 to 633)

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

Direct Submission  
 Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail: chimpan@gs.c.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/  
 Tel:81-45-503-9111, Fax:81-45-503-9170)  
 Clones are derived from the chimpanzee BAC library PTB. This BAC end  
 was generated during the R&D process and may have higher chance of  
 clone tracking errors.

## PRIMERS

## LIBRARY

## Vector

## R.Site 1

## R.Site 2

## Location/Qualifiers

## 1..633

## /organism="Pan troglodytes"

## /db\_xref="taxon:9598"

## /clone="PTB-080115.R"

## /sex="male"

## /cell\_type="lymphoblast"

## /clone\_1lb="PTB Chimpanzee Male BAC Library"

## BASE COUNT

## ORIGIN

## US-09-765-034-2 (1-334) x AG083174 (1-633)

OY 145 AlalIleThrVal-LeuValIleThrLeuGluLeuProIleLeuProLeuIleAsnProVa 164  
 DB 599 GCCATTGGGTTTATAGTAACATAGAGTCTACCTTACCTTAAAGAAATCCGTA 540  
 OY 164 ILeThrAspAsnGlyThrThrCysAsnAspPheAlaSerSerGlyAspProAsnTYrAs 184  
 DB 539 TAAGACTAACAGAGCGCCACCTTAATATTTTCAAGGTGTGAGACCCCAAAACAA 480

QY	184	nleuilelyrsertcysleuthleuenglypheuileiproleuphevalmecy	204
Db	479	CCTCATTTACAGACATGTGTCTACACCTGTGGGGTCTCTTATTCCTCTTTTGATGTG	420
QY	204	sphepetherytyrlyslleallaleupheuleylsglnargsanarglnalalathral	224
Db	419	TTTTCTTTTAATGCAAGATGGCTCTTCTCTGTAACAGAGAAATAGGAGGTGGCTGTGC	360
QY	224	aleuproleuglylulysproleuashnleuvalillemetkylalvalilliepheservalpr	244
Db	359	TCTATCCCTTGGAAAAGCCTTCACCTGATGTGTATCATATGACAGAGGTAACTTCTGTGCT	300
QY	244	ophethrprotyrhisvalmelargasnvalarglilelaserargleauglysertiply	264
Db	299	TTTTCACACCCCTTATCTGTCATCTGCGGAATGTGAGGTGCTCTCTCCCTGGGAGATTGGAA	240
QY	264	sglnatyrglncysthrnglnalvalilleasnserpetherylilevaljhrarproleual	284
Db	239	GCAGTNTCAAAAAGCAGGGAAATATCATCTCTTTACATTTGACACGCGCTTTGGC	180
QY	284	apheleuansservallleasnprovalpetherypheuleuglyasphilspheargas	304
Db	179	CTTTATGAAACAGTGCATCAACACCTGCTCTTATTTCTTTGGGAGATCATCTTCAGGGA	120
QY	304	pmetleuethsnnglnleuarghisasnphelysseriesleuthrserpheserargtpral	324
Db	119	CATGTGTATGATCAATCACTGACACACACTTCAAAATCCCTTACATCTTTACACATGAGC	60
QY	324	ahlsglu 326	
Db	59	TCTCGAG 53	
RESULT 3			
CNS04LVY3			
LOCUS		1101 bp	DNA
DEFINITION		Tetraodon nigroviridis genome survey sequence T3 end of clone	
		004B20 of library A from Tetraodon nigroviridis, genomic survey	
		sequence.	
ACCESSION		AL309576.1	GI:8216515
VERSION		AL309576.1	GI:8216515
KEYWORDS		GSS; genome survey sequence.	
SOURCE		Tetraodon nigroviridis.	
ORGANISM		Tetraodon nigroviridis.	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;	
		Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;	
		Tetraodontidae; Tetraodon.	
REFERENCE		1 (bases 1 to 1101)	
AUTHORS		Roeest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,	
		Bernot,A., Fitzames,C., Wincker,P., Brottier,P., Queller,F.,	
		Saurin,W. and Weissenbach,J.	
TITLE		Human gene number estimate provided by genome wide analysis using	
JOURNAL		Tetraodon nigroviridis DNA sequence	
REFERENCE		2 (bases 1 to 1101)	
AUTHORS		Roeest-Crollius,H., Jallion,O., Dasilva,C., Fitzames,C., Fisher,C.,	
		Bouneau,L., Billault,A., Queller,F., Saurin,W., Bernot,A. and	
		Weissenbach,J.	
TITLE		Characterization and repeat analysis of the compact genome of the	
JOURNAL		freshwater pufferfish Tetraodon nigroviridis	
REFERENCE		unpublished	
AUTHORS		3 (bases 1 to 1101)	
TITLE		Genoscope.	
JOURNAL		Direct Submission	
COMMENT		Submitted (12-APR-2000)	
		This sequence is a single read and was generated as part of a large	
		scale clone-and-sequencing project of the Tetraodon nigroviridis	
		genome. For more information, please take a look at	
		http://www.genoscope.cns.fr/Tetraodon.	
FEATURES		Location/Qualifiers	
SOURCE		1..1101	
		/organism="Tetraodon nigroviridis"	
		/db_xref="taxon:99883"	

		/clone="004B20"			
		/clone.lib="A"			
		/note="Genoscope sequence ID : COA004D0A10A2-end : T3			
BASE COUNT	225	a	319	c	274
ORIGIN			g	274	t
9 others					
Alignment Scores:					
Pred. No.:	5, 89e-67	Length:		1101	
Score:	692.00	Matches:		143	
Percent Similarity:	65.99%	Conservative:		53	
Best Local Similarity:	48.15%	Matches:		96	
Query Match:	39.61%	Indels:		5	
DB:	17	Gaps:		4	
US-09-765-034-2 (1-334) x CNS04VV3 (1-1101)					
QY	19	AlaAlaLeuGluGlySerTYRLeuSerIlePheTYRGIleGluPheValAlaGlyAl	38		
Db	79	GCTTACTGAGAGACTTCCTCCGCCACCTTACAGGTGTGAGTTCGCTGGTTT	138		
QY	39	LeuGlyAsnThrIleValTYRGIYTYRlePheSerLeuLysAsnTrpAsnSer	58		
Db	139	CCCCGGAACCTGGTGGTTCCTTGGTACATATTGGCTTGCACATGTGGCAGACTGC	198		
QY	59	AsnIleTYRleuPheAsnLeuSerValSerAspLeuAlaPheLeuGlyThrLeuProMet	78		
Db	199	CATATCAACCTCTTCAACCTGGCGCTCACAGCTTGTCTCTTCGACAGCTGCCGAC	258		
QY	79	LeuIleArgSerTYRAlaAsnGlyAsnTrpIleTYRGIYAspValLeuGlyIleSerAsn	98		
Db	259	CTCTCCACTCACTCAACCAACCAAGCTGCCTAAACAGTCCCTCTGTCATCGCCAAC	318		
QY	99	ArgTYRValLeuHisAlaAsnLeuTYRThrSerIleleuPheLeuThrPheIleSerIle	118		
Db	319	CGCTACATCCGCGACGCTCAACCTCTACTCTCATCTCTCATGTGGTGGCGAGCATG	378		
QY	119	AspArgTYRleuIleIleIleTYRTrpProPheArgGluHisLeuGluGlyIleGluPhe	138		
Db	379	GACCGCTACCTGCGTGGAGGACACCCGACCGGAACCACTGGCTGTGANC CGGGGCG	438		
QY	139	AlaIleLeuIleSerLeuAlaIleTrpValLeuValThrLeuGluLeuLeuProIleLeu	158		
Db	439	GGCGTGTGGTGGAGGGGCTCAGCTGTCTGTGTCAAGCTGGAGGTGGCCCCCATGATA	498		
QY	159	ProLeu---IleAsnProValIleThrAspAsnGlyTYRThrCysAsnAspPheAlaSer	177		
Db	499	GGCGTATGGTCCAGGACCTGACAGAGACAACTGGACACAGTGAAGCACTTCTCCAC	558		
QY	178	---SerGlyAspProAsnTYRAsnLeuIleTYRSetMetCysLeuThrLeuGluGlyPhe	196		
Db	559	CTGAGGAGAGACGTCACAC---AGCTTTGTTTACAGCTGGGAGTCACTCGTACGGGTAC	615		
QY	197	LeuIleProLeuPheValMetCysPhePheTYRTyrLysIleAlaLeuPheLeuGlyGln	216		
Db	616	TTCTCGCCCTCTGGGCGTGGGGTTTCACTTCCAGATCCGACCTGCTCCATGTC	675		
QY	217	ArgAsnArgGlnVal---AlaThrAlaLeuProLeuGluIleProLeuAsnLeuValIle	235		
Db	676	CAGGAGAGGCTTCGACGAGGCGACGCACTCTTCAAGGCGCCATCAAGGTGGT	735		
QY	236	MetAlaValValIlePheSerValProPheThrProTYRHisValMetArgAsnValArg	255		
Db	736	CTACGTGACGACATATTCCTGTCCTTACACCCCATACACAGCTGATGAGAAATCAAA	795		
QY	256	IleAlaSerArgLeuGlySerTrpLysGlnTYRGIYThrGlnValValIle-AsnSe	275		
Db	796	ATCCCTCTTACGAAGGTTTGGACCGAGTGGATTCATCAACGCTCGGATCAATCAAGGG	855		
QY	275	rPheTYRleValThrArgProLeuAlaPheLeuAsnSerValIleAsnProValPheTY	295		
Db	856	GCTGTACATTATTCACCCGACCTCTGGCGCTTTCGACAGGCTATTAACTCTCTTCA	915		
QY	295	rPheLeuLeuGlyAspHisPheArgAspMetLeuMetAsnGlnLeuArg	311		





[illegible]

[illegible]



JOURNAL  
COMMENT

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILNML, send email to:  
info@image.llnl.gov  
High quality sequence stop: 483.

FEATURES  
SOURCE

Location/Qualifiers  
1. 589  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3134907"  
/clone\_lib="NCI-CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP\_Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneids 1322376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 173 a 125 c 118 g 173 t  
ORIGIN

## Alignment Scores:

Pred. NO.: 2.63e-43 Length: 589  
Score: 478.00 Matches: 90  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 27.36% Indels: 0  
Gaps: 0

US-09-765-034-2 (1-334) x BF196066 (1-589)

OY 245 PheTTPProTyrHisValMetArgAsnValArgIleAlaSerArgLeuGlySerTrpIys 264  
|||||  
DB 589 TTTACACCCCTATGCATGCGGAATGAGATCGCTTCACGCGTGGGAGTTGGAAG 530  
OY 265 GlnTyrGlnCysThrGlnValAlaIleAsnSerPheTyrIleValThrArgProLeuAla 284  
|||||  
DB 529 CAGTATGATGACACGACGTCGTATCACTCTTTACATTGTGACACGGCCTTTGGCC 470  
OY 285 PheLeuAsnSerValIleAsnProValPheTyrPheLeuLeuGlyAspHisPheArgAsp 304  
|||||  
DB 469 TTTTGAACAGTGTATCAACCCCTGCTTCTTATTTCTTTGGGAGATCAGCTTCAGGAGAC 410  
OY 305 MetLeuMetAsnGlnLeuArgHisAsnPhelysSerLeuThrSerPheSerArgTrpAla 324  
|||||  
DB 409 ATGCTGATGAATCAACTAGACACAACTTCAATCCCTTACATCCTTGAAGATGGCT 350  
OY 325 HisGluLeuLeuSerPheArgGlyIys 334  
|||||  
DB 349 CATGAACCTCACTTCTTCATTCAGAGAAAAG 320

## RESULT 10

AM612141 639 bp mRNA linear EST 23-MAR-2000  
LOCUS AM612141/c  
DEFINITION hg94h07.x1 NCI CGAP Kid11 Homo sapiens cDNA clone IMAGE:2953309 3'

similar to SW:P2YR\_CHICK P34996 P2Y PURINOCEPTOR 1'', mRNA  
sequence.

ACCESSION AM612141  
VERSION AM612141.1 GI:7317327  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL  
COMMENT

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILNML at:  
image.llnl.gov/image/html/resources.shtml  
Seq primer: -400p from Glibco  
High quality sequence stop: 449.

FEATURES  
SOURCE

Location/Qualifiers  
1. 639  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2953309"  
/clone\_lib="NCI-CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP\_Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneids 1322376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 189 a 131 c 131 g 188 t  
ORIGIN

## Alignment Scores:

Pred. NO.: 3.85e-43 Length: 639  
Score: 477.00 Matches: 97  
Percent Similarity: 93.40% Conservative: 2  
Best Local Similarity: 91.51% Mismatches: 6  
Query Match: 27.30% Indels: 1  
DB: 10 Gaps: 0

US-09-765-034-2 (1-334) x AM612141 (1-639)

OY 230 ProLeuAsn-LeuValIleMetAlaValIlePheSerValProPheTrpProTyrHI 249  
|||||  
DB 634 CCTGCAATCTGTGGTCATCATGACGACGTGTAATATCTGTGCTATTAACACCATATCA 575  
OY 249 sValMetArgAsnValArgIleAlaSerArgLeuGlySerTrpIysGlnTyrGlnCysTh 269  
|||||  
DB 574 CGTCAATGCGATATGAGATGCGCTTCACGCTGGTGTGAAGGATGATCGATGAGTGCAC 515  
OY 269 rGlnValValIleAsnSerPheTyrIleValThrArgProLeuAlaPheLeuAsnSerVa 289  
|||||  
DB 514 TCAGTGTCTATCAACTCTTTACATTGTGACACGGCCTTTGGCCTTTGGAACAGTGT 455  
OY 289 IleAsnProValPheTyrPheLeuLeuGlyAspHisPheArgAspMetLeuMetAsnGI 309  
|||||  
DB 454 CATCAACCCCTGCTCTTCTTATTTCTTATGAGATCAGTTCAGGACATCGATGAATCA 395  
OY 309 nLeuArgHisAsnPhelysSerLeuThrSerPheSerArgTrpAlaHisGluLeuLeu 329  
|||||  
DB 394 ACTGAGACACAACTTCAATCCCTTACATCCTTGAAGATGGCTCATGAATCCTTACT 335  
OY 329 uSerPheArgGlyIys 334  
|||||  
DB 334 TTCAATTCAGAGAAAAG 319



RESULT 11  
BE221739/c 582 bp mRNA linear EST 03-JUL-2000  
LOCUS hr58c09.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:3132688 3'  
DEFINITION similar to SW:P2YR\_HUMAN P47900 p2y PURINOCCEPTOR 1 ;, mRNA  
sequence.  
ACCESSION BE221739.1 GI:8909046  
VERSION BE221739.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 582)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaaps-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov  
Seq primer: -40UP from Gibco  
High quality sequence stop: 479.  
FEATURES  
Source  
Location/Qualifiers  
1..582  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3132688"  
/clone\_1lb="NCI\_CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(clones 132376-132391, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 169 a 125 c 118 g 170 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 4,46e-42 Length: 582  
Score: 467.00 Matches: 88  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 26.73% Indels: 0  
DB: 10 Gaps: 0  
US-09-765-034-2 (1-334) x BE221739 (1-582)  
OY 247 ProTyHisValMetArgAnValArgilealaSerArgLeuGlySerTrpLysGlnTyr 266  
|||||  
DB 582 CCTATCAGCTCATGCGGAGATGTGAGGATCGCTTACACGCGCTGGGAGATTGAGCAGTAT 523  
OY 267 GlnCysThrGlnValValIleAsnSerPheTyrIleValIthrArgProLeuAlaPheLeu 286  
|||||  
DB 522 CAGTGGACATCAGGTGTGATCAACATCTCTTTTACATTGTGACAGGCGCTTTGCGCTTTCTG 463  
OY 287 AsnSerValIleAsnProValPheTyrPheLeuLeuGlyAspHisPheArgAspMetLeu 306  
|||||  
DB 462 AACAGTGTATCAACACCTGTCTTCTATTTCTTTTGGGAGATCAGCTTCAGGACATGCTG 403  
OY 307 MetAsnGlnLeuAlaArgHisAsnPheLysSerLeuThrSerPheSerArgTrpAlaHisGln 326

DB 402 ATGATCAACTGAGACACACTTCAATCCCTTACCTTTAGCAGATGGGCTCATGAA 343  
OY 327 LeuLeuLeuSerPheArgGlnLys 334  
|||||  
DB 342 CTCCTACTTTCATTCAGAGAAAG 319  
RESULT 12  
BB645274 636 bp mRNA linear EST 31-AUG-2001  
LOCUS BB645274  
DEFINITION Mus musculus cDNA clone B430012021 5', mRNA sequence.  
ACCESSION BB645274.1 GI:15402306  
VERSION BB645274.1  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 636)  
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Kono,H., Kouda,  
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,  
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takamashi,F., Takeda,Y., Tanaka,T., Toya,T.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resescgsc.riken.go.jp,  
URL:<http://genome.gsc.riken.go.jp/>  
Carninci,P., Shibata,Y., Hayatsu,M., Sugahara,Y., Shibata,K., Itoh,  
M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wag1,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Yamanaka,T., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa,  
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,  
Ishii,Y. and Hayashizaki,Y.  
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.  
Funct. Genomics 2 pre, L72-L86 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.  
e mouse tissues.  
FEATURES  
Source  
Location/Qualifiers  
1..636  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="B430012021"  
/clone\_1lb="RIKEN full-length enriched, 4 days neonate  
male adipose"  
/sex="male"  
/tissue\_type="adipose"  
/dev\_stage="4 days neonate"



Percent Similarity:	80.83%	Conservative:	12
Best Local Similarity:	70.83%	Mismatches:	22
Query Match:	23.01%	Indels:	3
DB:	10	Gaps:	0

US-09-765-034-2 (1-334) x BB846608 (1-416)

Oy	5	MetlaIrtrpAsnAlaItrhCysLysAsnTrpLeuAlaAlaGluAlaAlaLeuGluIuIyStyr	24
Db	59	ATGGCAGCAATTTATCTTGAGAAATGGTGGCAACCAAGCTATCTTGAATGAATGAC	118
Oy	25	TyrLeuSerIlePheTrpGlyIleGluPheValValGlyValLeuGlyAsnThrIleVal	44
Db	119	TACCTCTCGATTTTATGATGATGATGATTTTGGACTGCTTGGGAATGACATGCTG	178
Oy	45	ValTrpGlyTrpIlePheSerLeuLysAsnTrpAsnSerSerAsnIleTrpLeuPheAsn	64
Db	179	GTGTTGGCTACCTCTCTGATGAGAACGAGAAAGCAGCAATGCTATCTTTTAA	238
Oy	65	LeuSerValSerAspLeuAlaPheLeuCysThrLeuProMetLeuIleArgSerTrpAla	84
Db	239	CTTTCATC-TCGTGCTTGTCTTCTGTGCACCTTCATC-CTGATTAAGACTTATGCC	296
Oy	85	AsnGlyAsnTrpIleTrpGlyAspValLeuCysIleSerSnrArgTrpValLeuHISAla	104
Db	297	AATGATTAAGGGACCTATGAGATGTTCTATGATTAAGCAACGATATGGTTCACAA	356
Oy	105	AsnLeuTrpTrpSerIleLeuPheLeuThrPheIleSer-IleAspArgTrpLeuIle	123
Db	357	AACCTTAACACACATCTCTTCTCTCATTTATGACCATGACCGATATCTGCTC	414

RESULT 14  
AK017378  
LOCUS  
DEFINITION  
MUS MUSCULUS 6 DAYS NEONATE HEAD CDNA, RIKEN FULL-LENGTH ENRICHED  
LIBRARY, CLONE:5430432J15:purinergic receptor p2y, G-protein  
coupled 2, full insert sequence.  
ACCESSION  
AK017378  
VERSION  
AK017378.1  
KEYWORDS  
HTC: CAP TRAPPER.  
SOURCE  
MUS MUSCULUS (strain: C57BL/6J) 6 DAYS NEONATE HEAD CDNA TO mRNA,  
CLONE:5430432J15.  
ORGANISM  
MUS MUSCULUS  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE  
1  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
PUBMED  
99279253  
10349636  
AUTHORS  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
JOURNAL  
MEDLINE  
PUBMED  
20499374  
11042159  
REFERENCES  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohnaka, E., Watanabe, K.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
JOURNAL  
MEDLINE  
PUBMED  
11076661

REFERENCE	4
AUTHORS	Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Konno, H., Aachui, T., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Konno, S., Yamana, I., Saito, T., Okazaki, Y., Gojibori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, M., Gaasterland, T., Gissi, C., King, B., Koehn, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schirml, L. M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Pletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazza, R. L., Kombe, P., Nordone, P., Rang, B., Ringnadh, C., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohlsch, S., and Hayashizaki, Y. Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001) JOURNAL MEDLINE PUBMED 21085660 11217851 REFERENCE 5 (bases 1 to 2542) Aachui, T., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Aizawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanaoka, T., Hara, A., Hayatsu, N., Hill, D., Himoto, K., Hirooka, T., Horii, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schirml, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, Y., Teijima, Y., Toya, T., Yamamura, T., Yamana, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y. Direct Submission Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216) Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGATTCGAGTAAATTAATTAATCCCCCCCCC 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adaptor of sequence [5' GAGAGAGATTCGAGTAAATTAATTAATTAATCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified plasmidicp KS(+) after bulk excision from lambdaBda P1C I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B Location/Qualifiers 1. 2542 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="FANTOM.DB:5430432J15" /db_xref="MGI:1897121" /db_xref="taxon:10090" /clone="5430432J15" /tissue_type="head"

```

/clone.lib="RIKEN full-length enriched mouse cdna library"
/dev.strage="6 days neonate"
1..2542
/gene="p2ry2"
484..1605
/gene="p2ry2"
/notes="data source:MGD, source key:MGI:105107,
evidence:ISS
purinergic receptor P2Y, G-protein coupled 2
putative"
/codon_start=1
/protein_id="BAB30719.1"
/db_xref="GI:12856589"
/db_xref="MGI:105107"
/translation="MAADLEPMNSTINGTNGDELGKCRNEDFKYVLLPVSIGVVC
VLGLCNVALYIFLCRLKTNASTYTMFHLVADSDLSLAALPLVYVYAGDMPS
TVLCGLVFLFETNLYCSIFLTCISVIRCHISLWGLMGRVAVVWVL
VLACQAVLYEVTTSVGRGTCTCHDSARELSFHVASWMLDLFVPSVILVCY
VLACRLIKPKAYGTGGPLPRAKRSVRTIALVLAFCFLPHLITRLYSPSLDL
SCHTLNAINNAVYKTRPLASANSCLDPIVYLAQORLYVAFARADAKPRTEPPSQARR
KLGHRPRKRYKDLVSSDSRTESTPACSETKDRL"
BASE COUNT      517 a      737 c      705 g      582 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:      1.12e-31      Length:      2542
Score:          383.00      Matches:      96
Percent Similarity: 50.33%      Conservative: 58
Best Local Similarity: 31.37%      Mismatches: 136
Query Match:      21.92%      Indels:      16
DB:                11      Gaps:        7

US-09-765-034-2 (1-334) x AK017378 (1-2542)
Oy 7 TTPASNAATHrCysLys---AsnTRPLEuAlaAlaGluAlaLeuGlu----- 22
Db 505 TGGAAATACACCATGATGACGACCTGGAGGGGAGCAACCTGGATACAAAGTGTGCTTC 564
Oy 23 -----LysTyr---TyrLeuSerIlePheTyrGlyIleGluPheValValGly 37
Db 565 AAGCAGGACTTCAGATGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 624
Oy 38 ValLeuGlyAsnThrIleValValTyrGlyTyrIlePheSerLeuLysAsnTRPAsnSer 57
Db 625 TTGTGCCCGAAGCGTGTGCTCTATATCTCTATATGCGCGCTAAACCTGGAGACGC 684
Oy 58 SerAsnIleTyrLeuPheAsnLeuSerValSerAspLeuAlaPheLeuCysThrLeuPro 77
Db 685 TCCACACCATGATGATGCTGACGAGTTTCGACCTCTCTACGACGACGCTGCGCG 744
Oy 78 MetLeuIleArgSerTyrAlaAsnGly---AsnTRPLeuTyrGlyAspValLeuCysIle 96
Db 745 CTGTGGGTTTATTACTAGCCCGGGGCTGACCACTGGCCATTAGCAGCGCTGCTGCA 804
Oy 97 SerAsnArgTyrValLeuHisAlaAsnLeuTyrThrSerIleLeuPheLeuThrPheIle 116
Db 805 CTGGTCCCTTCCTCTTCTATACCAACCTCTAGTCGACATCTCTTCTCACCCTGCATC 864
Oy 117 SerIleAspArgTyrLeuIleIleLysTyrProPheArgGluHisLeuLeuGlnLys 136
Db 865 AGCGGACACCGGTGCTGCGAGTCTGCGCCCTGCACCTCCCTCGTGGGCGCGCGCC 924
Oy 137 GluPheAlaIleLeuIleSerLeuAlaIleTyrValLeuValThrLeuGluLeuLeuPro 156
Db 925 CGTATATGCCCCCGGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 984
Oy 157 IleLeuProLeuIleAsnProValIleThrAspAsnGlyThrThrCysAsnAspPheAla 176
Db 985 GTGCTTACTTCTGTCACACACGCGTGGGGAACCCGGATCACCCTGCAATGACCTCG 1044
Oy 177 SerSerGlyAspProAsnTyrAsnLeuIleTyrSerMetCysLeuThrLeuLeuGlyPhe 196
Db 1045 GCCGAGAGAGCTTTTACGCAATTTTGTGCTTACACGCTCCGTCATGCTGCGTCTTTT 1104

```

```

Oy 197 LeuIleProLeuPheValMetCysPhePheTyrTyrLysIleAlaLeuPheLeuLysGln 216
Db 1105 GCTGTGCCCTTTTCCGTAATCCTGTGCTGTATACGCTTATAGCCAGCGGCTGCTCAA 1164
Oy 217 ArgAsnArgGluValAlaThrAlaLeuPro-----LeuGluLysProLeuAsnLeuVal 234
Db 1165 CCGGCTTATGGACACACAGAGCTGTGCTGCGGCCAACCCGAATCTGTGGCACCATT 1224
Oy 235 IleMetAlaValAlaIlePheSerValProPheThrProTyrHisValMetArgAsnVal 254
Db 1225 GCGTTAGTATGCGCGCTTGTGCGCCCTGCTGCTGCTTCCACCTTCACGGACACCTTC 1284
Oy 255 ArgIleAlaSerArgLeuGlySerTyrLysGlnTyrGlnCys---ThrGlnValValIle 273
Db 1285 TACTACTCTCTCCGATCA-----CTTGACCTCAGCTGCACACCCCTCAACGCCATC 1335
Oy 274 AsnSerPheTyrIleValThrArgProLeuAlaPheLeuAsnSerValIleAsnProVal 293
Db 1336 AACATGGCATATATAGATACCCGCGCTGCGACGCCACCAACAGTTGCTTACCCGGTA 1395
Oy 294 PheTyrPheLeuLeuGly 299
Db 1396 CTCTACTCTCTGCGCAGGG 1413

```

```

RESULT 15
AK005013
LOCUS
DEFINITION
Mus musculus adult male liver cDNA, RIKEN full-length enriched
library, clone:1300015C04;purinergic receptor P2Y, G-protein
coupled 2, full insert sequence.
ACCESSION
AK005013.1 GI:1283638
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA,
clone.lib:RIKEN full-length enriched mouse cdna library
clone:1300015C04.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
PUBMED
10349636
2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
PUBMED
11042159
3 Shibata,K., Itoh,M., Aizawa,K., Nagoka,S., Sasaki,N., Carninci,P.,
Kono,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishi,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,
Yamamoto,R., Matsuno,T., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matshiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunaga,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
PUBMED
11076861
4 Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
Aizawa,T., Hara,A., Fukunishi,Y., Kono,H., Adachi,J., Fukuda,S.,
Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamana,K.I.,
Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,

```



